

ABSTRACT OF THE DISCLOSURE

A document is compared to the documents in a document collection using a hash
algorithm and collection statistics to detect if the document is similar to any of the documents in
5 the document collection.

DC2-228176.3

00000000000000000000000000000000